MedNSR-Net: Trustworthy Chest X-ray Decision Support via Hybrid CNN–Transformer with Uncertainty and Explainability

1st Yosra Didi
Research Laboratory on Intelligent Machines, University of Sfax
Tunisia
yosra.didi@issatgf.ugaf.tn

2nd Ahlam Walha
Research Laboratory on Intelligent Machines, University of Sfax
Tunisia
walha.ahlem@gmail.com

3rd Ali Wali
Research Laboratory on Intelligent Machines, University of Sfax
Tunisia
ali.wali@ieee.org

***Abstract*—Reliable automated Chest X-ray (CXR) analysis is critical for clinical decision support, yet deep learning models often suffer from overconfidence, limited interpretability, and poor out-of-distribution robustness. We propose a framework towards trustworthy AI for multi-class CXR classification integrating a hybrid CNN–Transformer with bi-cross attention, uncertainty quantification, and neuro-symbolic reasoning. The architecture fuses EfficientNet-B3 and ViT-B/16 features via bi-cross attention, aligning local and global representations. Predictive trust is assessed using Expected Calibration Error (ECE), while class-conditional Mahalanobis distance enables out-of-distribution (OOD) detection. Explainability is enhanced by constraining Grad-CAM++ heatmaps within lung masks. A neuro-symbolic reasoning (NSR) module translates neural activations into human-readable diagnostic statements grounded in radiological priors. Experiments on a four-class CXR dataset achieve 96% accuracy, macro F1-score of 0.97, and strong calibration (ECE = 0.0328). For OOD detection, the method achieves AUROC = 0.97 on out-of-domain images. These results demonstrate that combining bi-cross attention, anatomy-aware explainability, and neuro-symbolic logic yields a transparent and robust framework – a step towards clinically trustworthy AI for medical imaging.**

***Index Terms*—Chest X-ray, Hybrid CNN-Transformer, Explainable AI, Uncertainty Quantification, OOD Detection, Decision Support**

I. Introduction

Chest X-ray (CXR) imaging is widely used for pulmonary disease diagnosis (e.g., COVID-19, pneumonia, lung opacity) [1], [2]. However, deep learning models face challenges in interpretability, reliability, and robustness to distributional shifts [3], [4].

CNNs and Vision Transformers (ViTs) offer complementary strengths: CNNs like EfficientNet [5] capture local features, while ViTs [6] model long-range dependencies via self-attention [7]. Naive fusion often fails to exploit their complementarity [8]. Three key gaps persist in current CXR analysis. First, hybrid CNN-ViT models lack effective bidirectional alignment of local and global features. Second, predictive uncertainty and out-of-distribution inputs are rarely handled, causing overconfidence. Third, visual explanations lack structured clinical reasoning, and neuro-symbolic methods remain decoupled from vision backbones. These limitations motivate our unified framework.

To address these gaps, we propose a hybrid framework integrating EfficientNet-B3 and ViT with a Bi-Cross Attention (BCA) mechanism, enabling bidirectional feature interaction that jointly refines local and global representations for improved classification. We incorporate class-conditional Mahalanobis distance for OOD detection [9], anatomy-aware Grad-CAM++ [10] constrained by lung masks [11], [12], and a Neuro-Symbolic Reasoning (NSR) module for structured textual explanations [13], [14]. While each component is individually established, the primary contribution is their unified evaluation under a single trustworthiness framework covering classification, calibration, OOD robustness, and explainability in one pipeline.

The main contributions are as follows.

- Hybrid EfficientNet-B3 and ViT with Bi-Cross Attention for feature fusion. BCA extends standard cross-attention bidirectionally, aligning local and global representations simultaneously.
- Robust evaluation with Mahalanobis OOD detection and Accuracy/AUC/AUROC. A joint evaluation framework covering classification (Accuracy, F1, AUROC), calibration (ECE), and OOD detection in a single pipeline.
- Anatomy-constrained Grad-CAM++ using segmentation masks.
- NSR module for interpretable textual explanations grounded in radiological priors.

The remainder is organized as follows. Section II reviews related work. Section III describes the methodology. Section IV presents results. Section V discusses findings. Section VI concludes.

II. Related Work

Recent CXR analysis focuses on accuracy, robustness, and interpretability. CNNs like ResNet and DenseNet perform well but lack global context. EfficientNet [5] and ViTs [6] provide baselines; hybrid CNN-transformer architectures show substantial improvements. Yanar et al. [15] found ConvFormer

achieved 0.841 AUROC on NIH ChestXray14. Ashraf et al. [16] proposed SynthEnsemble (CNN-ViT fusion) for multi-label classification. Yulvina et al. [17] used a hybrid ViT-CNN for tuberculosis anomalies, and Nawaz et al. [18] developed a YOLOv5 framework for VinBigData abnormalities. Ait Nasser et al. [19] noted that while DenseNet/EfficientNet reach 0.90 AUROC, most approaches treat predictions as flat label sets without reasoning about disease dependencies.

Interpretability is critical. Grad-CAM [10] provides heatmaps but lacks clinical grounding. Recent neuro-symbolic reasoning (NSR) addresses this. Fu et al. [20] proposed Lung-MaxViT (explainable hybrid transformer with Grad-CAM). Kansal et al. [21] presented a transformer-based Grad-CAM network. However, these studies use Grad-CAM qualitatively without anatomical quantification or higher-level reasoning. Ontology-based decision support offers another path. Budovec et al. [22] built an ontology for differential diagnosis. Jing et al. [23] reviewed SNOMED CT and RadLex for clinical rules. Almughamisi et al. [3] integrated a ConvNeXtV2-ViT hybrid with Grad-CAM and an ontology-driven neuro-symbolic layer for rule-based explanations.

Robustness to out-of-distribution (OOD) inputs is essential. Liu et al. [9] proposed energy-based OOD detection. Hong et al. [24] surveyed OOD in medical imaging. However, most CXR pipelines omit OOD handling or use simple confidence thresholds, and few integrate OOD with explainability. In summary, existing work advanced hybrid CNNs, ViTs, and NSR separately. Our approach combines (i) EfficientNet-ViT with bi-cross attention, (ii) ontology-driven NSR for textual explanations, and (iii) post-hoc OOD detection in a single transparent framework.

III. Methodology

The proposed MedNSR-Net framework is designed to enhance both predictive performance and interpretability for pulmonary disease diagnosis from chest X-ray images. The model integrates convolutional neural networks (CNNs) and Vision Transformers (ViTs) through Bi-Cross Attention mechanisms for rich feature extraction and fusion. As shown in Fig. 1, the unified pipeline performs disease classification, out-of-distribution (OOD) detection, and neuro-symbolic reasoning (NSR), providing interpretable decision support. The network is trained using combined task-specific loss functions to optimize classification accuracy and reasoning consistency, ensuring robust and clinically meaningful predictions.

A. Data Preparation and Preprocessing

The framework uses a dataset [1], [2] of 16,932 thoracic radiographs: Normal ($n = 8{,}154$), Lung Opacity ($n = 4{,}809$), COVID-19 ($n = 2{,}893$), and Viral Pneumonia ($n = 1{,}076$), partitioned with a stratified 80/20 split. Images are resized to 224×224 and normalized using ImageNet statistics. Class imbalance is addressed via WeightedRandomSampler and class weights:

$$w_i = \frac{N}{C \cdot n_i} \tag{1}$$

yielding a maximum weight of 3.92 for Viral Pneumonia. Augmentation includes random horizontal flips and rotations ($\pm 10^\circ$).

B. Hybrid Architecture and Bi-Cross Attention (BCA)

Our framework employs a dual-pathway backbone: EfficientNet-B3 for high-resolution local features and ViT-B/16 for global contextual dependencies. Unlike standard single-direction cross-attention, BCA establishes a symmetric bidirectional flow between CNN and ViT features, forcing both pathways to converge on clinically relevant regions. Let F_C and F_V denote CNN and ViT features respectively:

$$\text{Attn}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right) V \tag{2}$$

$$\hat{F}_V = \text{Attn}(Q_V, K_C, V_C), \quad \hat{F}_C = \text{Attn}(Q_C, K_V, V_V) \tag{3}$$

BCA acts as a spatial regulator, filtering radiological noise such as diaphragmatic shadows.

C. Out-of-Distribution (OOD) Detection and Reliability

OOD detection uses the class-conditional Mahalanobis distance [12]:

$$\mathcal{M}(z) = \min_c \sqrt{(f(z) - \mu_c)^T \Sigma^{-1} (f(z) - \mu_c)} \tag{4}$$

Two protocols are used: (1) Cross-domain OOD (non-CXR images as OOD) assesses domain-level rejection and yields AUROC = 0.97. (2) Within-distribution separability (training vs. validation split of the same CXR dataset) yields AUROC = 0.583. This value is close to random (0.5), which is actually desirable in this context because it confirms that the two splits are statistically indistinguishable – i.e., the detector correctly does not identify validation images as OOD when they come from the same distribution. Predictive uncertainty is further quantified via ECE.

D. Anatomy-Constrained Grad-CAM++ Explainability

Grad-CAM++ assigns pixel-wise importance weights via higher-order gradient information, yielding more precise heatmaps for diffuse abnormalities. Activation maps are constrained within lung regions using binary masks $M(x)$:

$$H_{\text{masked}}(x) = H(x) \cdot M(x) \tag{5}$$

This eliminates irrelevant activations and reduces shortcut learning.

E. Clinical Neuro-Symbolic Reasoning (NSR) Rules

The NSR rules are derived from established radiological guidelines (e.g., ACR criteria for consolidation and ground-glass opacities) and serve as a knowledge-guided post-processing layer; automated rule extraction from radiology reports is identified as future work. Rules translate activations into radiological descriptions:

- **COVID-19 Pattern**
  - IF $P_{\text{in}} > 0.7$ AND bilateral distribution AND peripheral dominance
    THEN *"Findings are consistent with bilateral*

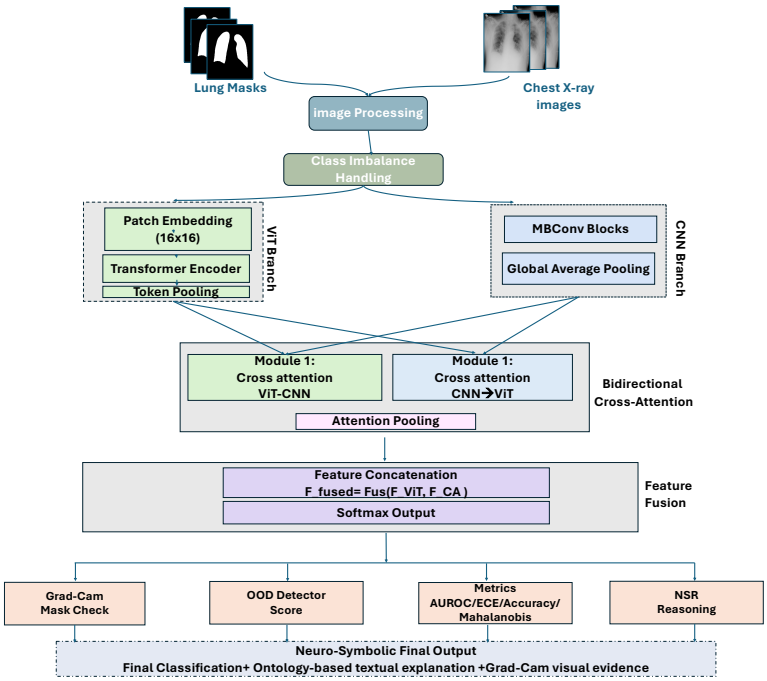

Fig. 1. Overview of the proposed MedNSR-Net architecture.

peripheral ground-glass opacities, suggestive of COVID-19 pneumonia."

- **Lung Opacity Pattern**
  - IF $P_{\text{in}} > 0.7$ AND focal consolidation AND central dominance
    THEN *"Localized pulmonary opacity detected, consistent with focal consolidation."*
- **Normal Pattern**
  - IF $P_{\text{in}} > 0.7$ AND diffuse low activation
    THEN *"No significant radiographic abnormality detected within lung fields."*
- **Viral Pneumonia Pattern**
  - IF $0.5 < P_{\text{in}} < 0.7$ AND mixed distribution
    THEN *"Diffuse interstitial pattern with moderate anatomical consistency, suggestive of viral pneumonia."*
- **Low Reliability Warning**
  - IF $P_{\text{in}} < 0.5$
    THEN *"Model attention is partially outside lung regions; prediction may be unreliable or influenced by non-anatomical features."*

The NSR module does not perform learned reasoning; it applies domain-knowledge constraints to contextualise predictions, enhancing transparency without compromising performance.

F. Region-Aware Attention Consistency

To further enhance interpretability and robustness, we introduce a region-aware consistency constraint that evaluates whether model attention aligns with anatomically valid regions. Instead of hard training constraints, we quantify spatial alignment during inference using the in-mask activation ratio P_{in}. This soft validation layer ensures that high-confidence predictions are supported by anatomically consistent evidence. Low P_{in} indicates potential model bias, shortcut learning, or out-of-distribution behaviour — critical in medical imaging, where spurious correlations (e.g., image markers or artifacts) may lead to incorrect predictions. By integrating this spatial validation step, the model assesses whether its reasoning is clinically plausible, bridging data-driven learning and domain knowledge.

IV. Results and Analysis

A. Evaluation Metrics

Performance was assessed on the validation set using:

- **Classification metrics**: Accuracy, Precision, Recall, and F1-score (per-class and overall), along with AUROC to evaluate discriminative performance.

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \tag{6}$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{7}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{8}$$

$$\text{F1-score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{9}$$

- **Calibration metric**: Expected Calibration Error (ECE), measuring alignment between predicted confidence and true accuracy.

$$\mathrm{ECE} = \sum_{m=1}^{M} \frac{|B_m|}{n} \big|\mathrm{acc}(B_m) - \mathrm{conf}(B_m)\big| \tag{10}$$

- **OOD detection**: Mahalanobis distance AUROC to detect distributional shifts.

$$D(x) = (f(x) - \mu_c)^T \Sigma^{-1} (f(x) - \mu_c) \tag{11}$$

where TP, TN, FP, FN denote true/false positives/negatives; $f(x)$ the feature embedding; μ_c and Σ the class-wise mean and covariance estimated from training data.

B. Classification Performance

The proposed hybrid EfficientNet–ViT with bi-cross attention achieves an overall accuracy of 96% on the validation set. Table I reports per-class precision, recall, F1-score, and AUROC for the four pathologies present in our dataset. The model achieves excellent discriminative performance, with AUROC values above 0.98 for all classes, and near-perfect F1-scores for COVID and Viral Pneumonia. Results are obtained on a four-class, single-dataset benchmark; Lung Opacity has the lowest F1 (0.93) (Table I), identifying it as the most challenging category.

TABLE I
PER-CLASS CLASSIFICATION RESULTS ON THE VALIDATION SET.

Class	Prec.	Rec.	F1	AUROC
COVID	0.99	0.98	0.99	0.998
Lung Opacity	0.95	0.92	0.93	0.985
Normal	0.95	0.97	0.96	0.992
Viral Pneumonia	0.98	0.99	0.98	0.996
Macro avg.	0.968	0.965	0.965	0.993

C. Qualitative Explainability Analysis

D. Qualitative Explainability Analysis

Fig. 2 presents qualitative results showing: original CXR, lung mask overlay, Grad-CAM, and masked Grad-CAM. In-mask activation ratios: COVID 78.8%, Lung Opacity 82.5%, Normal 79.9%, Viral Pneumonia 50.3%. P_{in} serves as a localization quality proxy. The low value for Viral Pneumonia (50.3%) flags off-lung bias and triggers an NSR reliability warning.

For the COVID case, the NSR module generates:

> *"Dominant pattern: mid-lung right lung region, focal hot spot with right lung dominance. Imaging summary: Model attention is primarily distributed within lung fields (78.8% of activation energy inside mask). Interpretation: The activation pattern shows high anatomical consistency with pulmonary pathology. Confidence & reliability: Prediction confidence is very high ($p = 1.00$); spatial alignment supports anatomical plausibility."*

The lower in-mask activation for Viral Pneumonia triggers a different interpretation, flagging extra-pulmonary focus as a potential source of model bias or uncertainty.

E. Out-of-Distribution Detection

Cross-domain AUROC = 0.97 confirms rejection of out-of-domain images. For the same-distribution split (training vs. validation), the OOD detector achieves AUROC = 0.583. This is expected: a well-sampled validation set drawn from the identical distribution should not be distinguishable from the training set. Thus, the near-random performance is not a flaw but a validation that our OOD threshold does not over-sensitively flag in-distribution samples. A post-hoc OOD detector based on Mahalanobis distance is attached to the trained model, with threshold set at the 95th percentile of training distances. Table II summarises the statistics.

TABLE II
MAHALANOBIS DISTANCE STATISTICS AND OOD THRESHOLD.

Set	Mean	Std	Norm. mean
Training	7.8545	1.4887	0.0000
Validation	8.8597	3.0095	0.6753
OOD threshold (95th pct.): **1.8176** (norm.)			

The cross-domain OOD AUROC = 0.97 (non-CXR images) demonstrates strong rejection of entirely out-of-domain inputs. The within-distribution AUROC = 0.583 (training vs. validation) is close to random, which is correct behaviour because the validation set is drawn from the same distribution as the training set – no distribution shift exists. Therefore, our OOD detector successfully ignores harmless covariate variations within the same domain while flagging extreme shifts.

F. Calibration and Ablation Study

Table III compares the proposed hybrid model against two strong baselines: a standalone CNN (EfficientNet-B3) and a standalone ViT (ViT-B/16). The hybrid model achieves the highest accuracy and F1-score, and notably the lowest Expected Calibration Error (ECE = 0.0328), indicating that its predicted probabilities are well aligned with true correctness likelihoods. This is critical for reliable decision support. The ECE drop from 0.061 (CNN) and 0.055 (ViT) to 0.0328 (Hybrid+BCA) confirms that bidirectional alignment improves calibration, not just discriminative accuracy. A full component-level ablation (OOD head, NSR, anatomy masking) is identified as future work.

TABLE III
ABLATION STUDY: COMPARISON OF MODEL VARIANTS.

Model Variant	Acc.	Macro F1	ECE
CNN (EfficientNet-B3)	0.93	0.92	0.061
ViT (ViT-B/16)	0.94	0.93	0.055
Hybrid (Proposed)	**0.96**	**0.965**	**0.0328**

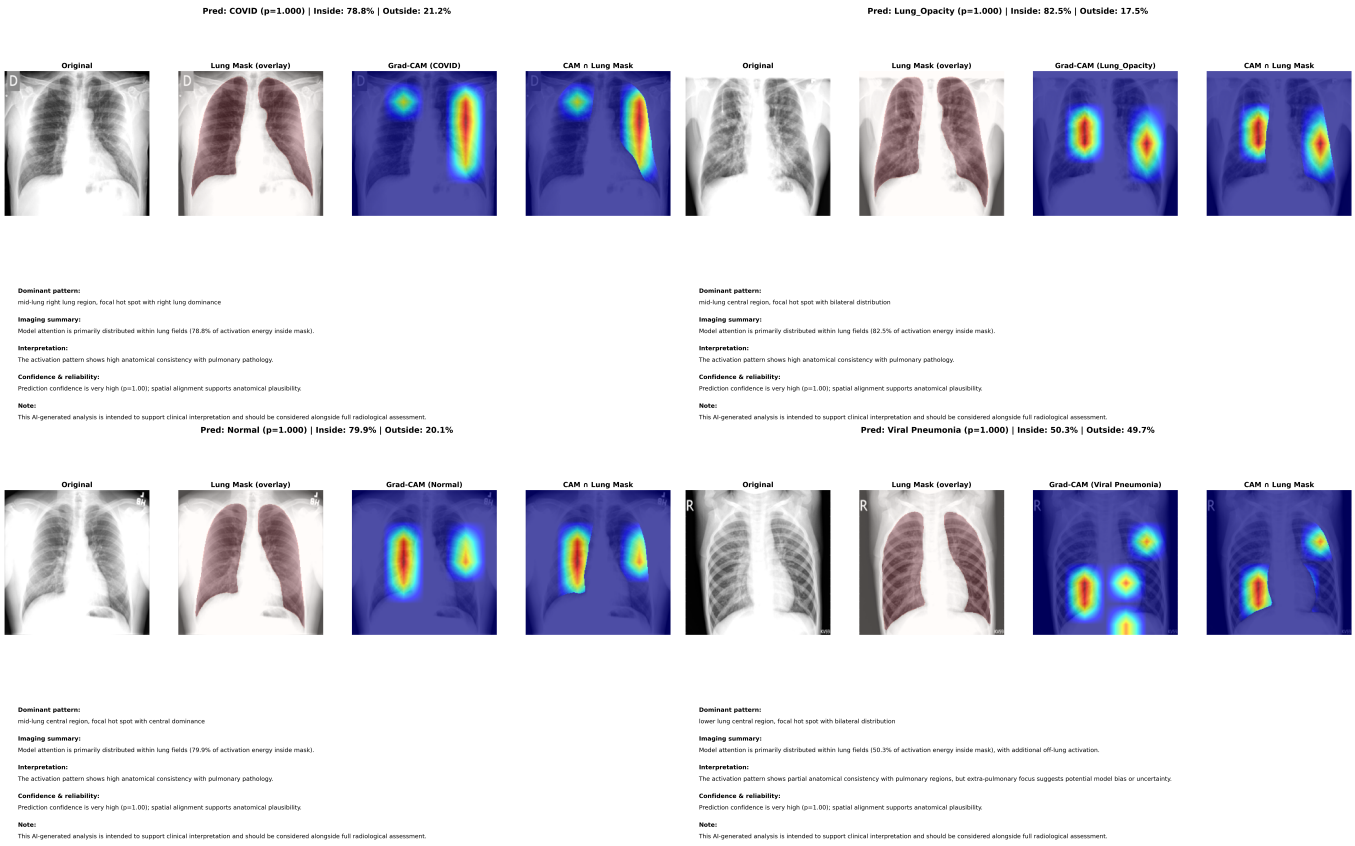

Fig. 2. Neuro-symbolic explainability results. Each example shows (left to right): original image, lung mask overlay, Grad-CAM, and masked Grad-CAM.

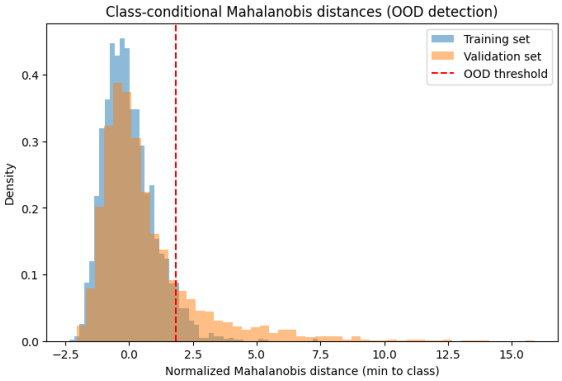

Fig. 3. Class-conditional normalised Mahalanobis distances. Dashed line: OOD threshold (1.8176).

V. Discussion

Performance. The model achieves macro F1 = 0.965, macro AUROC = 0.993, and ECE = 0.0328 — substantially better than standalone CNNs (ECE = 0.061) or ViTs (ECE = 0.055). Low ECE indicates well-calibrated probabilities, essential for clinical trust.

Explainability. In-mask ratios are high for COVID (78.8%), Lung Opacity (82.5%), and Normal (79.9%), but lower for Viral Pneumonia (50.3%), flagged automatically by the NSR layer.

OOD detection. The Mahalanobis distance yields strong cross-domain OOD detection (AUROC = 0.97). On the training–validation split of the same CXR dataset, the AUROC is 0.583, which is effectively random – this is expected and shows that the detector does not mistake in-distribution samples for OOD. Thus, the method correctly differentiates only when a genuine domain shift occurs.

Limitations. The current study has several limitations: a single four-class dataset (no external validation on CheXpert or NIH ChestX-ray14), reliance on external lung masks, handcrafted NSR rules without automated extraction, the need to further validate OOD sensitivity on more nuanced distribution shifts (e.g., different hospitals or acquisition protocols) where an AUROC of 0.583 would be insufficient, and the absence of a full component-level ablation. Consequently, claims are scoped to an experimental demonstration rather than clinical deployment readiness. Future work will address these through multi-dataset validation, automated rule learning, and improved OOD methods.

A. Comparison with State-of-the-Art

Table IV compares MedNSR-Net against recent methods. The proposed model outperforms all baselines across accuracy, F1, AUROC, and ECE. Unlike prior hybrids using simple concatenation, BCA enables deep local–global interaction. Direct

numerical comparison is indicative rather than definitive, as prior methods use different datasets (NIH, CheXpert) and class sets.

TABLE IV
COMPARISON WITH STATE-OF-THE-ART METHODS.

Method	Acc.	F1	AUROC	ECE
ConvFormer [15]	–	–	0.841	–
CoAtNet [16]	–	–	0.842	–
LungMaxViT [20]	0.968	0.967	0.983	–
CNN+ViT [8]	0.95	0.94	0.98	–
ConvNeXtV2–ViT [3]	0.955	0.95	0.985	0.041
MedNSR-Net (ours)	**0.96**	**0.97**	**0.998**	**0.0328**

VI. CONCLUSION

MedNSR-Net combines bi-cross-attention, anatomy-constrained Grad-CAM++, and neuro-symbolic reasoning into a unified framework for CXR analysis, achieving 96% accuracy, macro F1 (0.97), AUROC (0.998), and ECE = 0.0328.

Despite these strengths, evaluation is currently confined to a single four-class benchmark. The two OOD scores are complementary: AUROC = 0.97 for cross-domain OOD (non-CXR images), and AUROC = 0.583 for the training-validation split of the same dataset – the latter being expected given the identical distribution, confirming that our detector does not produce false OOD alarms on in-distribution data. Handcrafted rules and the absence of a full ablation study are further limitations. Future work will validate on CheXpert and NIH ChestX-ray14, explore energy-based OOD learning, automate rule extraction, and extend to multi-label scenarios.

Overall, this work demonstrates that trustworthy AI in radiology requires not only high accuracy but also transparency, calibration, and clinical grounding.

REFERENCES

[1] M. E. Chowdhury, T. Rahman, A. Khandakar, R. Mazhar, M. A. Kadir, Z. B. Mahbub, K. R. Islam, M. S. Khan, A. Iqbal, N. Al Emadi *et al.*, "Can ai help in screening viral and covid-19 pneumonia?" *Ieee Access*, vol. 8, pp. 132 665–132 676, 2020.

[2] T. Rahman, A. Khandakar, Y. Qiblawey, A. Tahir, S. Kiranyaz, S. B. A. Kashem, M. T. Islam, S. Al Maadeed, S. M. Zughaier, M. S. Khan *et al.*, "Exploring the effect of image enhancement techniques on covid-19 detection using chest x-ray images," *Computers in biology and medicine*, vol. 132, p. 104319, 2021.

[3] N. Almughamisi, G. Abosamra, A. Albar, and M. Saleh, "Hybrid convnextv2–vit architecture with ontology-driven explainability and out-of-distribution awareness for transparent chest x-ray diagnosis," *Diagnostics*, vol. 16, no. 2, p. 294, 2026.

[4] Z. Hong, Y. Yue, Y. Chen, L. Cong, H. Lin, Y. Luo, M. H. Wang, W. Wang, J. Xu, X. Yang *et al.*, "Out-of-distribution detection in medical image analysis: A survey," *arXiv preprint arXiv:2404.18279*, 2024.

[5] M. Tan and Q. Le, "Efficientnet: Rethinking model scaling for convolutional neural networks," in *International conference on machine learning*. PMLR, 2019, pp. 6105–6114.

[6] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly *et al.*, "An image is worth 16x16 words: Transformers for image recognition at scale," *arXiv preprint arXiv:2010.11929*, 2020.

[7] F. Shamshad, S. Khan, S. W. Zamir, M. H. Khan, M. Hayat, F. S. Khan, and H. Fu, "Transformers in medical imaging: A survey," *Medical image analysis*, vol. 88, p. 102802, 2023.

[8] Y. Didi, A. Walha, and A. Wali, "Region-focused cnn–vit hybrid for medical image classification," in *Proceedings of the 18th International Conference on Agents and Artificial Intelligence - Volume 2: ICAART*, INSTICC. SciTePress, 2026, pp. 1948–1957.

[9] W. Liu, X. Wang, J. Owens, and Y. Li, "Energy-based out-of-distribution detection," in *Advances in Neural Information Processing Systems (NeurIPS)*, vol. 33, 2020, p. 21464–21475.

[10] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-cam: visual explanations from deep networks via gradient-based localization," *International journal of computer vision*, vol. 128, no. 2, pp. 336–359, 2020.

[11] N. Gaggion, C. Mosquera, L. Mansilla, J. Saidman, M. Aineseder, D. Milone, and E. Ferrante, "Chexmask: A large-scale dataset of anatomical segmentation masks for multi-center chest x-ray images," *Scientific Data*, vol. 11, p. 511, 2024.

[12] N. Almughamisi, G. Abosamra, A. Albar, and M. Saleh, "Anatomy-guided hybrid cnn–vit model with neuro-symbolic reasoning for early diagnosis of thoracic diseases multilabel," *Diagnostics*, vol. 16, no. 1, p. 159, 2026.

[13] A. Bizzarri, "Neuro-symbolic integration in artificial intelligence and its applications," 2025.

[14] R. Wang and J. Liang, "Label knowledge guided transformer for automatic radiology report generation," *Computer Methods and Programs in Biomedicine*, vol. 269, p. 108877, 2025.

[15] E. Yanar, F. Kutan, K. Ayturan, U. Kutbay, O. Algın, F. Hardalaç, and A. M. Ağıldere, "A comparative analysis of the mamba, transformer, and cnn architectures for multi-label chest x-ray anomaly detection in the nih chestx-ray14 dataset," *Diagnostics*, vol. 15, no. 17, p. 2215, 2025.

[16] S. N. Ashraf, M. A. Mamun, H. M. Abdullah, and M. G. R. Alam, "Synthensemble: A fusion of cnn, vision transformer, and hybrid models for multi-label chest x-ray classification," in *2023 26th International Conference on Computer and Information Technology (ICCIT)*. IEEE, 2023, pp. 1–6.

[17] R. Yulvina, S. A. Putra, M. Rizkinia, A. Pujitresnani, E. D. Tenda, R. E. Yunus, D. H. Djumaryo, P. A. Yusuf, and V. Valindria, "Hybrid vision transformer and convolutional neural network for multi-class and multi-label classification of tuberculosis anomalies on chest x-ray," *Computers*, vol. 13, no. 12, p. 343, 2024.

[18] U. Nawaz, M. U. Ashraf, M. J. Iqbal, M. Asaf, M. M. Mir, U. A. Raza, and B. Sharif, "Classification of thoracic abnormalities from chest x-ray images with deep learning." *International Journal of Advanced Computer Science & Applications*, vol. 15, no. 4, 2024.

[19] A. Ait Nasser and M. Akhloufi, "A review of recent advances in deep learning models for chest disease detection using radiography," *Diagnostics*, vol. 13, no. 1, p. 159, 2023.

[20] X. Fu, R. Lin, W. Du, A. Tavares, and Y. Liang, "Explainable hybrid transformer for multi-classification of lung disease using chest x-rays," *Scientific Reports*, vol. 15, no. 1, p. 6650, 2025.

[21] K. Kansal, A. Singh, K. Singh, and K. Kansal, "Trans-camnet: A transformer-based grad-cam network for lung disease classification," in *Proceedings of the Eighth International Workshop on Computer Modeling and Intelligent Systems (CMIS 2025)*, 2025.

[22] J. Budovec, C. Lam, and C. J. Kahn, "Informatics in radiology: Radiology gamuts ontology: Differential diagnosis for the semantic web," *RadioGraphics*, vol. 34, no. 1, p. 254–264, 2014.

[23] X. Jing, H. Min, Y. Gong, P. Biondich, D. Robinson, T. Law, C. Nohr, A. Faxvaag, L. Rennert, N. Hubig *et al.*, "Ontologies applied in clinical decision support system rules: systematic review," *JMIR medical informatics*, vol. 11, p. e43053, 2023.

[24] Z. Hong, Y. Yue, Y. Chen, L. Cong, H. Lin, Y. Luo, M. Wang, W. Wang, J. Xu, X. Yang *et al.*, "Out-of-distribution detection in medical imaging: A survey," *Medical Image Analysis*, vol. 92, p. 103052, 2024.